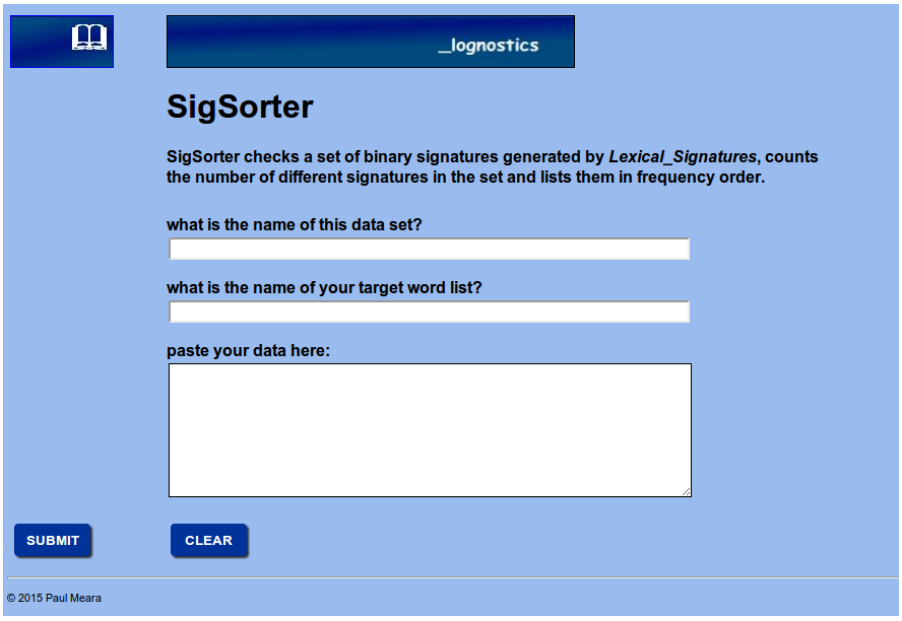
SigSorter: The Manual
Paul Meara

SigSorter is a companion program to V_LexSig. It helps you perform some simple counts and sorting operations on sets of texts described by Lexical Signatures. Therefore, you will need the output produced by V_LexSig as input for SigSorter.

1. The SigSorter workspace looks like this:



2. **Data preparation**: SigSorter works with a text file that contains a set of report strings generated by the V_LexSig program described in the previous section. The data should look something like this:

1101111100/TList12/ThreeBears01.txt
1110111100/TList12/ThreeBears02.txt
1100011100/TList12/ThreeBears03.txt
1100011100/TList12/ThreeBears04.txt
1101111100/TList12/ThreeBears05.txt
1100011100/TList12/ThreeBears06.txt
1100111101/TList12/ThreeBears07.txt
1100111111/TList12/ThreeBears08.txt
1100111110/TList12/ThreeBears09.txt
1100111111/TList12/ThreeBears10.txt

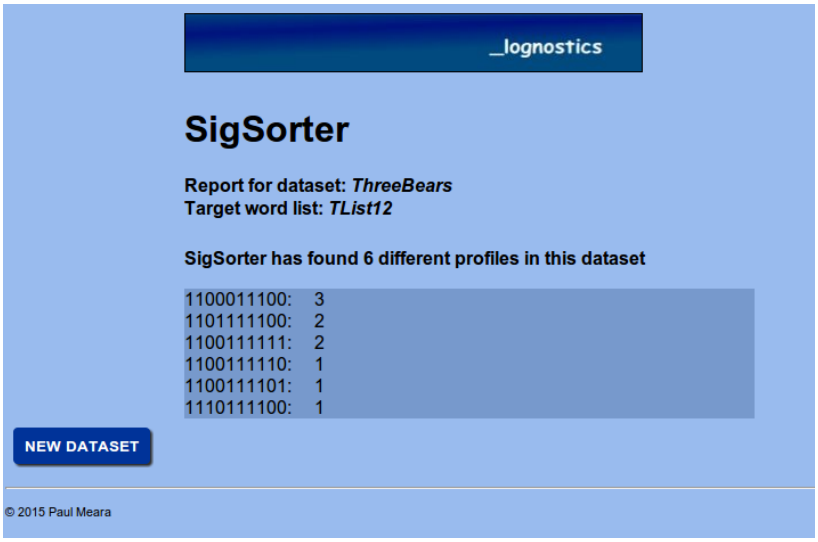
Each line consists of a report string comprised of three parts separated by a slash (/). The first part contains a binary description of the relevant text. The second part consists of the name of the Target Word List that was used to generate the report string. The third part is a string which identifies the text that was being assessed. For example, in the list above, we have ten texts that were all assessed using Target Word List *TList12*. This list contained ten target words, so the binary descriptions contain ten digits.

Normally, you will be using report strings that have been prepared using the V_LexSig program, but you can prepare your data by hand if you need to do so.

3. Enter a name for the data set that you want to analyse.
Then enter the name of the Target Word List that you used to generate the original data.
Finally, copy and paste your data set into the large box on the workspace.

4. Click the **Submit** button to process this data.

5. The SigSorter report screen is shown below:



This report page tells us that SigSorter has processed a data set called *ThreeBears,* and that this file was constructed using Target Word List *TList12*. The report also tells us that six different signatures were identified in the data set, and these signatures are listed in order of their frequency in the data set. In this report, one signature occurred three times, two signatures occurred twice and there were three unique signatures that occurred only once in the data set.

Copy and paste this signature list to a text file if you want to save it for further analysis.

6. Click on the **New dataset** button if you want to analyse another set of texts, or if you want to work with a new Target Word List.

Background reading:

Meara, P.M., Jacobs, G. and Rodgers, C. (2002) Lexical Signatures in foreign language free-form texts. *ITL Review of Applied Linguistics* 135-136, 85-96.

Meara, P.M., Rodgers, C. and Jacobs, G. (2000) Computational assessment of texts written by L2 speakers. *System* 28 (3), 345-354.

For a more detailed discussion of *SigSorter*, see **PM Meara and I Miralpeix (2017)** *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.